

Государственный архив Свердловской области (ГАСО). Ф. Р-2033. Свердловское областное управление профессионально-технического образования.

Разинков С. Л., Захаровский Л. В. От источник-ориентированной базы данных к Digital Humanities: опыт цифровизации архивной информации по истории профтехобразования // Управление документами в цифровой экономике: проблемы взаимодействия : материалы II Междунар. науч.-практ. конф. (Москва, 6 ноября 2019 г.) / под общ. ред. А. Б. Безбородова. М. : РГГУ, 2019. С. 288–297.

Российский государственный архив экономики (РГАЭ). Ф. 4372. Государственный плановый комитет Совета Министров СССР (Госплан СССР).

Центр документации общественных организаций (ЦДООСО). Ф. 4. Свердловский областной комитет КПСС.

УДК 004.658.2

Т. И. Славко

## АРХИВЫ И БАЗЫ ДАННЫХ В ЦИФРОВОЙ ЭКОНОМИКЕ: ПРОБЛЕМЫ КОМПЛЕКТОВАНИЯ И ИСПОЛЬЗОВАНИЯ

В работе рассматривается системный подход к формулированию и использованию баз данных в современных условиях развития информационного общества. Особо значимым информационным ресурсом становится документированная информация на цифровых носителях. Перед документоведами стоит важнейшая проблема создания и ведения систем электронных документов, перед архивистами - проблемы архивного хранения и использования таких систем. Начало создания баз данных в исторической науке связано с решением источниковедческих проблем анализа массовых источников. Этот опыт важен и в современных условиях.

Проблемы комплектования, выявления, хранения и анализа баз данных, как целостного ресурса, базируются на принципах теории больших систем, а также методах и технологиях анализа больших данных. Поэтому в сложной системе баз данных целесообразно выделение нескольких подсистем, взаимосвязанных между собой. Это могут быть три подсистемы, каждая из которых, по мере необходимости, подразделяется на более удобные. Первая подсистема – совокупность баз данных с документированной информацией, вторая – содержит методы анализа баз данных, третья – научно-справочный аппарат.

*Ключевые слова:* цифровая экономика, базы данных, структурированная и неструктурированная информация, сложные документационные системы, анализ данных.

Развитие современного информационного общества, связанное с цифровизацией управленческой и экономической деятельности, массовым внедрением электронных документов (ЭД) и электронного документооборота, анализом данных с помощью разнообразных информационных технологий ставят перед документоведами и архивистами множество научных проблем, как теоретического, так и прикладного характера. От их успешного решения зависят темпы реализации многих стратегических задач, сформулированных в федеральных законах и государственных программах.

В условиях цифровизации стоит вопрос обработки больших объемов данных, которые представляют «совокупность подходов, инструментов и методов автоматической обработки структурированной и неструктурированной информации, поступающей из большого количества различных, в том числе

разрозненных или слабосвязанных источников информации, в объемах, которые невозможно обработать вручную за разумное время» [О стратегии развития информационного общества].

Цифровизация потребовала и обсуждения такого актуального вопроса как систематизация данных, что связан с эффективным поиском необходимых ресурсов в информационном пространстве. В качестве информационного ресурса подразумевается «совокупность систематизированной информации с определенной структурой информационного ресурса, порядком его формирования, ведения, изменения, а также применения содержащейся в нем информации» [О систематизации и гармонизации информации...].

Важнейшим информационным ресурсом в данном случае является документированная информация на цифровых носителях. Исходя из этого, перед документоведами стоит важнейшая проблема формирования и ведения систем ЭД в различных сферах деятельности, а перед архивистами – проблемы архивного хранения и использования таких систем. Остановимся на рассмотрении некоторых из этих проблем, в основном в плане постановки и актуализации задач.

Современные информационные ресурсы могут быть представлены в различных формах, в том числе и в виде специализированных баз данных (БД), содержащих сведения об однородных явлениях и процессах.

Начало создания БД в исторической науке с конца 1950-х годов связано с введением в научный оборот массовых источников и использованием для их анализа математических методов. Под массовыми здесь подразумеваются структурированные данные, отражающие множество однородных объектов наблюдения, позволяющих раскрыть те или иные тенденции и закономерности, присущие изучаемому явлению или процессу.

К массовым источникам относятся статистические материалы первичного или сводного характера, например, первичные бланки разнообразных переписей, отложившиеся в архивах.

К массовым источникам причисляются также документы, имеющие стандартные формы, близкие или идентичные к статистическим. Это совокупность анкетных данных на членов трудовых коллективов, отдельных социальных групп. Массовыми являются и документы, которые, хотя и не имеют разработанных форм, но описывают стандартные ситуации (биографии, характеристики, отчеты и т. д.). Нарративные источники, содержащие развернутые индивидуальные тексты, после классификационных преобразований с помощью так называемого контент-анализа, составляли наиболее обширную группу массовых источников.

Фактически, по каждому из массовых источников, речь шла о создании реляционной базы данных, которая затем обрабатывалась с помощью ЭВМ на уровне группировки, средних показателей, коэффициентов взаимосвязей и др.

Для формирования БД проводилось докомпьютерное структурирование данных в традиционной форме (на бумажных носителях) с помощью предварительно разработанной анкеты. В ней фиксировался определенный набор признаков, ответы на которые содержались во всех или в большинстве единиц наблюдения источника. Для каждого признака предлагалась и единая система измерения.

Реляционная БД представляет собой двумерные таблицы, в структуре которой есть поля (столбцы с названием признака), характеризующего все однородные объекты и записи (каждая из которых соответствует одной единице наблюдения).

Как правило, создаваемые БД были нацелены на реализацию исследовательских задач. Они позволяли ввести новые данные, обработка которых ручным способом была крайне затруднительна. Именно БД расширяли возможности использования различных методов исследования – структурного и системного анализа, моделирования и др.

Процесс создания БД был связан с выявлением массовых источников, их источниковедческой критикой, разработкой методики измерения признаков, кодировкой, написанием и использованием программного аппарата и т. д. Такие БД предполагали возможность многоразового использования. Однако чаще всего они были нацелены на решение ограниченного числа исследовательских задач и не давали возможность для вторичного обращения другим исследователям.

Как правило, материалы БД откладывались в личных архивах их создателей, не поступая на государственное или хоть сколько-то централизованное хранение. В лучшем случае, публиковались результаты обработки, а в государственные архивы передавались первичные записи на бумажных носителях-перфокартах.

Централизованное хранение БД осложнялось несколькими причинами. Прежде всего, это постоянная модификация вычислительной техники и языков программирования. Массовое использование информационных технологий в прикладных целях требовало от гуманитариев специальных знаний в области математических методов обработки данных и компьютерных технологий. Отсутствовали стандарты описания БД для пользователей документоведов, историков, архивистов и других специалистов гуманитарных направлений. Все это сдерживало не только процесс создания новых БД, но и решение задач их архивного хранения и неоднократного использования.

В связи с массовым переходом на электронные документы и электронный документооборот роль БД, содержащих документированную информацию, резко повышается. Формируются БД, обеспечивающие деятельность органов государственного и муниципального управления, создаются БД для принятия эффективных управленческих решений. Множество БД нацелено

на получение и распространение новых знаний в различных сферах жизни общества.

Концептуальный подход к систематизации БД с целью их комплектования, выявления, хранения и использования должен базироваться на принципах теории больших систем, появление и развитие которой датируется началом 1970-х годов. Исследование больших систем связано с изучением сложных систем на уровне их структуры, моделей закономерностей с помощью математических методов анализа. Современные БД включают как структурированную, так и неструктурированную информацию.

В этом случае в сложной системе БД целесообразно выделить несколько подсистем, взаимосвязанных между собой, что облегчит доступ к их информации. Считаем возможным в этом случае выделить три подсистемы: первая – совокупность БД с документированной информацией; вторая – содержащая методы анализа БД; третья – научно-справочная. Каждая из перечисленных подсистем, в свою очередь, может быть представлена более дробными структурными элементами.

Первая подсистема может быть классифицирована по видам и разновидностям ЭД, по характеру реализации управленческих функций и т. д. Подсистема, связанная с методами анализа данных, в качестве структурных элементов для реляционных БД включает систему операций, базирующихся на языке реляционной алгебры. Их, как известно, семь: объединение БД, пересечение, разность, произведение, выборка, проекция, соединение.

Для БД, содержащих неструктурированные данные, имеющие свободную форму представления информации, следует предусмотреть новые методы и технологии анализа, активно разрабатываемые в последние два десятилетия в рамках теории больших данных.

Структурными элементами второй подсистемы БД является система математико-статистических методов, методов теории графов, имитационного моделирования, теории игр, линейного программирования, теории массового обслуживания и др.

Третья подсистема, связанная с научно-справочным аппаратом, предполагает описание каждой БД с точки зрения характеристики информационной структуры ЭД и их свойств. В качестве справочного аппарата язык описания БД должен предназначаться для пользователей – не только для IT-специалистов, но и гуманитариев.

Итак, формирование разнообразных по назначению БД с помощью современных информационных и коммуникационных технологий связано с решением ряда теоретических и прикладных проблем. Одной из них является систематизация БД с точки зрения их комплектования, хранения и использования. Крайне важной при этом является стандартизация в рамках техноло-

гии Больших информационных систем, сосредоточенных на единой платформе. Основная роль в решении таких проблем принадлежит документоведам и архивистам.

#### Список литературы

О систематизации и гармонизации информатизации в Российской Федерации: проект Федерального закона // Минфин России [сайт]. URL: [https://minfin.gov.ru/common/upload/npa\\_projects/Poyasnitelnaya\\_zapiska\\_530.docx](https://minfin.gov.ru/common/upload/npa_projects/Poyasnitelnaya_zapiska_530.docx) (дата обращения: 09.07.2020).

О стратегии развития информационного общества в Российской Федерации на 2017–2030 годы : Указ Президента РФ от 09.05.2017 №203 // Президент России [сайт]. URL: <http://kremlin.ru/acts/bank/41919> (дата обращения: 09.07.2020).

УДК 94(470.5).06.004.65

С. И. Цеменкова

### КАРТОГРАФИЧЕСКИЕ ДОКУМЕНТЫ УРАЛЬСКОЙ ГОРНОЙ АДМИНИСТРАЦИИ ПЕРВОЙ ПОЛОВИНЫ XVIII века: ОПЫТ СОЗДАНИЯ БАЗЫ ДАННЫХ<sup>1</sup>

В статье представлены результаты работы по формированию базы данных, содержащей сведения о картографических документах, созданных в период активного промышленного освоения Урала в первой половине XVIII в. горной администраций Екатеринбурга. Характеризуются сведения о структуре базы данных, ее содержании и назначении.

*Ключевые слова:* база данных, картография, картографический документ, история Урала, первая половина XVIII в.

Электронная база данных «Картографические документы Уральской горной администрации первой половины XVIII века» разработана в рамках научной деятельности коллектива Лаборатории эдиционной археографии Уральского гуманитарного института Уральского федерального университета.

Основу базы данных составляют результаты анализа и формализации сведений о составе и содержании картографических документов, созданных в деятельности уральской горной администрации в первой половине XVIII в. В основу базы легли документы, извлеченные из фонда № 59 «Чертежная Уральского горного управления» Государственного архива Свердловской области. Документы фонда учтены в 17 описях, включающих в себя более 30 тыс. дел за период с начала XVIII по начало XX в. [Краткий справочник..., с. 32].

Интерес к картографическим документам первой половины XVIII в. возник не случайно. Это период активного промышленного освоения Урала, и

---

<sup>1</sup> Статья подготовлена при поддержке Российского научного фонда № 20-18-00233 «Екатеринбург в 1733 г.: историко-антропологическая и архитектурно-пространственная реконструкция».